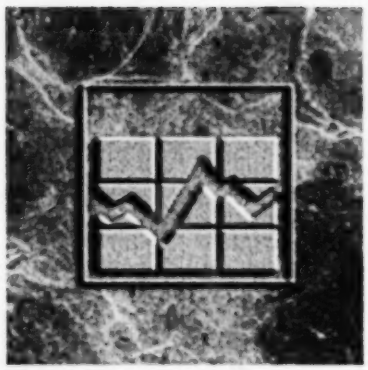
Catalogue no. 62F0026MIE — No. 004
ISSN: 1708-8879
ISBN: 0-662-40938-8

Research Paper

Household expenditures research paper series

Methodology of the Food Expenditure Survey

2001



by Sylvie Auger, Khushnood Kha.. .. d Johanne Tremblay

Household Survey Methods Division
R.H. Coats Building, Ottawa, K1A 0T6

Telephone: 613 951-7355

Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-7355); (888) 297-7355 income@statcan.ca.

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 62F0026MIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Income Statistics Division

Household expenditures research paper series

Methodology of the Food Expenditure Survey

2001

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2005

All rights reserved. The content of this publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it is done solely for the purposes of private study, research, criticism, review, newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, name of product, catalogue, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopy, for any purposes, without the prior written permission of Licensing Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

July 2005

Catalogue no. 62F0026MIE2005004

ISSN: 1708-8879
ISBN: 0-662-40938-8

Frequency: Occasional

Ottawa

La version française de cette publication est disponible sur demande (n° 62F0026MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Abstract

The Food Expenditure Survey (FES) is a periodic survey collecting data from households on food spending habits. Data are collected mainly using weekly diaries of purchases that the respondents must fill in daily during two consecutive weeks.

This paper presents a detailed description of the methodology of this survey. First, we briefly described the sample design which is mainly based on the plan of the Labour Force Survey. Then we present the methods of collection, data processing, weighting, and variance estimation, as well as the suppression of unreliable data in the tables of estimates.

Table of contents

1. Introduction **6**

2. Target population **6**

3. Sample design **7**

3.1 Size and allocation of the FES sample 8
3.2 LFS sample design (cluster selection) 9
3.3 Selection of the FES sample 11

4. Data collection **11**

4.1 Data collection method 12
4.2 Interview and follow-up procedures 12
4.3 Monitoring and controls 13
4.4 Non-response to the FES 13

5. Data processing **14**

5.1 Coding, data capture and editing 14
5.2 Imputation 15

6. Weighting and estimation **16**

6.1 Design weight 17
6.2 Adjustment for total non-response 17
6.3 Adjustment to auxiliary data 18
6.4 Other adjustments 19
6.5 Estimates 20

7. Estimation of the sampling error **20**

7.1 Model for deriving an approximation of the CV for domain estimates 21
7.2 Model for deriving an approximation of the CV from microdata files 21

8. Suppression of data and confidentiality **21**

8.1 Suppression of unreliable data in the tables of estimates 21
8.2 Confidentiality of the microdata files 22

9. Changes in the survey methodology **22**

Bibliography **23**

Appendix A **24**

Formula for calculating the variance of estimates using the jackknife method

Appendix B **25**

Formula for approximating the CV for a domain (a sub-group of the population)

1. Introduction

The Food Expenditure Survey (FES) is a periodic survey conducted every four to six years. The survey collects data from Canadian households on food spending habits.

This survey is used to complete the Survey of Household Spending (SHS)[1] by gathering detailed data on food expenditures that cannot be collected using the SHS methodology. The Food Expenditure Survey and the Survey of Household Spending are used to update the weighting used by the Consumer Price Index.

The FES data are gathered through an in-person interview using a paper questionnaire (FE2) and through two weekly diaries of purchases (FE3) that the respondent must fill in daily during two consecutive weeks. The data were gathered from a sample of households covering the 10 provinces as well as the cities of Whitehorse and Yellowknife. In 2001, the city of Iqaluit was added. The survey is conducted every month during a single calendar year. The collected data are then edited and weighted. The survey's products include the publication, tables and microdata files required by various users.

This paper provides a detailed description of the survey's methodology relating to the sample design, data collection and processing, production of estimates and other products, as well as the rules governing their dissemination.

2. Target population

The target population of the FES includes individuals residing in private Canadian households, but excludes official representatives of foreign countries living in Canada and their families, and individuals living on Indian reserves and crown lands. The restriction to private households means that people living full-time in institutions such as prisons, chronic care hospitals and residences for senior citizens, as well as members of religious and other communal colonies, members of the Armed Forces living in military camps, and individuals living full-time in hotels or rooming houses are excluded.

1. The SHS is an annual survey in which a sample of Canadian households is asked to report on all expenditures made during a calendar year. The results of the SHS are published in *Spending Patterns in Canada*, Catalogue No. 62-202.

We did not collect data from persons temporarily living away from their family (for example, students at university) because the data would be gathered from their families if selected in the sample. Therefore, if a household is made up only of persons temporarily living away from their family, it is considered an out of scope household.

Furthermore, in 2001, since the sample size is much smaller than in previous surveys, members of households residing in dwellings located in remote areas are excluded from the survey. These areas, which are sparsely populated in general, are located mainly in the northern parts of certain provinces. These regions are difficult to access and therefore data collection for these households is very costly.

The survey therefore covers about 97% of the population in the 10 provinces. In the Yukon, Northwest Territories and Nunavut, the survey covers only the cities of Whitehorse, Yellowknife and Iqaluit, which represent 62%, 44% and 17% of their respective populations based on the 1996 Census.

3. Sample design

Food expenditure data are collected from households coming from a sample of dwellings selected according to a multi-stage, stratified sampling design. This design varies according to the level of urbanization but generally consists of a two-stage sample for which the first level is an area sample, i.e., a sample of geographic areas called clusters.

The selection of clusters is done for the complete year. This set of selected clusters is divided evenly into 12 subsets for data collection to be done during the 12 months of the year — that is, one subset per month. The collection is done during the whole year so as to take into account the seasonality of food expenditures. The method of dividing the 12 subsets must ensure that the three monthly samples of each quarter cover Canada geographically.

In the second stage, before each month of collection, dwellings are selected from a list of all private dwellings from the subset for that particular month. All the selected dwellings that are inhabited by individuals from the target population constitute the survey sample.

To reduce the collection costs, the FES uses largely the same sample design as the Labour Force Survey (LFS). The dwellings in the FES sample are selected from LFS sample clusters, but the two surveys use different dwelling samples. The main aspects of the LFS cluster sample design are described in Section 3.2. A more detailed description is available in the LFS methodology publication [1]. The approach used in selecting FES dwellings from LFS sample clusters is covered in Section 3.3. The following section explains how the sample is allocated among the five Canadian regions and the three cities of the territories.

3.1 Size and allocation of the FES sample

Due to budget constraints, the sample size of the 2001 survey has been reduced considerably compared to the previous survey of 1996. During the 1996 survey, the sample size was 17,000 dwellings resulting in about 15,000 households after excluding vacant dwellings and out of scope households (e.g., the households made up only of individuals having a permanent residence elsewhere). With this sample size, it was expected that the estimates at the national and provincial levels as well those for the cities of the territories[2] would be reliable.

At the planning stage of the FES 2001, it was decided to reduce the sample size from 17,000 dwellings to 14,000 dwellings. Even with this reduction, relatively reliable national estimates, provincial estimates, and estimates for the three cities of the territories were still expected. A few weeks after the start of data collection, in January 2001, the sample size had to be reduced from 14,000 to 9,000 dwellings due to budget constraints. Since collection was already in progress, it was not possible to reduce the January sample. The sample was reduced appropriately starting from the month of February. Reliable estimates were now expected to be produced at the regional level[3] instead of the provincial level.

The total sample, excluding the sample from the three cities of the territories, was allocated among the five Canadian regions so as to obtain a compromise between national and regional reliability of *Total weekly food expenditures*. More specifically, the sample allocation is based on the variability of food expenditure in each region and, to a lesser extent, the household counts at the regional level.

In each region, the sample was distributed to provinces proportionally by their household counts. Each provincial sample was distributed in direct proportion to the size of the population of the census metropolitan areas. The portion of the sample allocated outside the census metropolitan areas matched the LFS allocation [1].

Table 3.1 presents the sample size for each of the provinces and the three cities of the territories for the 1996 and 2001 surveys. The sample size is presented in terms of number of dwellings and number of households (after excluding vacant dwellings and out of scope households).

---

2. In the territories, only the cities of Yellowknife, Whitehorse and Iqaluit are covered by the survey.
3. The five regions are Atlantic, Quebec, Ontario, Prairies and British Columbia.

Table 3.1
Sample size by province and three cities of the territories

Geography	Sample size			
	1996		2001	
	Number of Dwellings	Number of Households	Number of Dwellings	Number of Households
Canada	**17,000**	**14,993**	**9,488**	**8,414**
Newfoundland and Labrador	691	592	233	199
Prince Edward Island	621	516	79	67
Nova Scotia	1,347	1,143	396	329
New Brunswick	1,124	941	304	261
Atlantic Region	**3,783**	**3,192**	**1,012**	**856**
Quebec Region	**2,395**	**2,139**	**1,816**	**1,615**
Ontario Region	**3,886**	**3,479**	**2,632**	**2,404**
Manitoba	1,023	895	451	387
Saskatchewan	1,369	1,211	393	335
Alberta	1,461	1,319	948	868
Prairies Region	**3,853**	**3,425**	**1,792**	**1,590**
British Columbia Region	**2,212**	**2,038**	**1,558**	**1,422**
Total regions	**16,129**	**14,273**	**8,810**	**7,887**
Whitehorse	435	363	250	202
Yellowknife	436	357	252	192
Iqaluit	0	0	176	133
Total 3 cities of the territories	**871**	**720**	**678**	**527**

3.2 LFS sample design (cluster selection)

The LFS sample design is based on data from the Census of Population and is redesigned after each decennial census to take into account the changes in the population. The current design is based on 1991 Census data.

The principles underlying the LFS sample design are the same for every province. First, each province is divided into a number of geographic regions, made up of Economic Regions (ERs), Employment Insurance Economic Regions (EIERs) and of their intersections. In particular, every census metropolitan area forms a geographic region since it is an EIER.

Each geographic region is then divided into types of areas, primarily urban areas, rural areas and remote areas. The sample design varies according to the type of area. Remote areas represent the northern parts of some provinces and are, for the most part, sparsely populated. These areas are excluded from the FES. The sample design of the urban and rural areas is described below.

Urban areas

In some major cities with large numbers of apartment buildings, both an apartment list frame and an area frame are used. In other urban areas, only an area frame is employed.

An area frame is made up of a list of geographic zones. These zones are combined to form strata. There can be up to three levels of stratification. At the top levels, the aim is generally to form geographically compact and contiguous strata, whereas at the bottom level the requirement is for final strata to be as homogeneous as possible with respect to certain socio-economic characteristics. In a few large cities[4], separate strata are formed from enumeration areas with high average household income (about $100,000 or more).

To reduce collection costs, the dwellings that make up the final stratum are not selected directly. The stratum is first divided into clusters. In urban areas, the clusters may be combinations of block-faces, enumeration areas (EAs) or parts of EAs. Then clusters are selected (usually six, sometimes a multiple of six) in each stratum with a probability proportional to cluster size. For example, if the size of one cluster is twice as large as the one of another cluster, the former will be twice as likely to be selected as the latter.

The apartment list frame is a list of apartments prepared using information from the Canada Mortgage and Housing Corporation. This frame provides better representation of apartment residents and minimizes the effect of cluster growth due to construction of new apartment buildings. In some cities[5], apartment strata are divided into two categories: low-income (where the average household income is under $20,000) and regular. For each stratum in the frame, apartment buildings are selected for the first-stage sample with a probability proportional to the number of apartments in the building.

In low-density urban areas, which are highly dispersed towns, a different sample design is used. Sampling is done in three stages: first, towns are selected within the strata; next, clusters (block-faces) are chosen within the towns; finally, dwellings are selected within the clusters.

Rural areas

Only an area frame is used in rural areas. Geographic strata are formed by combining two or three census divisions, which are then subdivided, where numbers permit, to form strata that are homogeneous with respect to certain socio-economic characteristics. In the first stage of sampling, enumeration areas are selected within each final stratum with a probability proportional to the number of dwellings in the EA.

In low-density rural areas, a variation on the sample design is used. Two or three primary sampling units each consisting of a group of six EAs are selected in the

4. Montreal, Ottawa, Toronto, Hamilton, London, Winnipeg, Calgary and Vancouver.
5. Montreal, Ottawa-Hull, Toronto, Winnipeg, Calgary, Edmonton and Vancouver.

first stage, and then a sample of dwellings is selected within each of the EAs in the selected primary sampling units.

3.3 Selection of the FES sample

Interviewers visit the clusters selected in the LFS sample design and make a list of all private dwellings they contain. From that list, one sample is chosen for the LFS and a different one is selected for the FES. Dwellings are selected by systematic sampling.

Since the FES uses a much smaller sample than the LFS, dwellings are not selected in every LFS cluster. The LFS is a panel survey in which households remain in the sample for six months. The LFS sample was designed so that it could be divided into six representative subsamples to permit rotation of one sixth of the sample each month. That is why six clusters (or a multiple of six) are selected in each final stratum, one per rotation group. This method makes it easy to select a smaller sample for another survey since a subset of the rotation groups can be used. This is generally the approach used for LFS supplementary surveys. For the FES, the number of rotation groups is determined at the stratum level according to the survey's specific needs with regard to provincial and subprovincial allocation of the sample. In some instances, only part of a rotation group is needed. Where this is the case, dwellings are removed from the sample with a systematic drop on a geographically ordered list.

The dwelling sample is obtained following the cluster listing operation. Since the sampling rates are predetermined, there may be a difference between the expected and actual sample sizes if the number of dwellings on the list differs from the number used in developing the survey's sample design. To keep collection costs in check (since cluster sizes tend to increase) and prevent significant disparities in interviewer workloads, two methods are used to control the sample size.

The first method, which is used in general, consists to correct the problem by systematically removing some of the originally selected dwellings. This process of keeping the sample size at the desired level is known as sample stabilization. The second method is used when the number of dwellings increases sharply in certain urban areas. In this case, cluster subsampling is used instead. There are three options, depending on how large the increase is and how similar the new dwellings are to others in the same stratum: form subclusters, create a new stratum, or subsample the dwellings in the cluster.

4. Data collection

The FES gathers information on food expenditure by Canadian households on a voluntary basis.

4.1 Data collection method

Data are collected through in-person interviews using a paper questionnaire (FE2) and two weekly diaries of purchases (FE3) that the respondent must fill out daily over two consecutive weeks. The survey is conducted every month during a single calendar year and the data are collected from households coming from a different sample of dwellings each month.

The questionnaire is used primarily to gather data on certain socio-economic characteristics (for example, number of people, household income) and information on household purchasing habits and their food expenditures during the previous month.

The diaries serve to collect daily food expenditures in stores and in restaurants. In the case of spending done in stores, the details include a description of the item purchased, the type of processing (fresh, frozen, canned, dried, other), the type of store (supermarket, specialty store, corner store, other), the number of units purchased, the net weight or volume per unit (in metric or imperial measure) and the total cost of the purchase. In the case of spending done in restaurants, the details include the type of meal (breakfast, lunch, dinner, snack), the number of meals purchased, the type of restaurant (table service, fast food, cafeteria, other), and the total cost of the purchase, excluding the costs, taxes and tips related to alcoholic beverages.

In addition, for 2001, respondents were strongly encouraged to attach their receipts to the page in the diary corresponding to the date of the purchase. There was another new element in 2001, namely the notebook. All members of the household aged 10 years and older were given a small notebook to carry around at all times in which to record products at the time of purchase.

Members of the household

To obtain the expenditures of a household, it is essential to properly identify the members that make up that household. The person or group of people who occupy a dwelling constitute a household. For the FES, the members of that household are the persons living in the dwelling at the time of the interview who do not have a permanent residence elsewhere and who are not members of another household.

4.2 Interview and follow-up procedures

The interviews are conducted by Statistics Canada interviewers, some of whom also collect LFS data. These interviewers receive training specific to the FES.

One or two weeks before the month of the interview, the regional office mails an introductory letter to the occupants of the selected dwellings stressing the importance of the survey and the confidentiality of the information gathered, along with a brochure providing additional information on the survey. The interviewer then goes to the dwelling. If no one is home, further attempts are made to contact the household: for example, by visiting at different times of the day or by telephoning. The telephone number, if available, is part of the

information on the dwelling which is given to the interviewer. If the telephone number is missing, the interviewer tries to track it down.

During the initial meeting, the interviewer himself fills out the questionnaire (FE2) by recording the answers obtained from the respondent. He then gives the diaries (FE3) to the respondent providing explanations on how to complete them properly and setting a time to come back eight days later to collect the completed diary. At the time of the second visit, the interviewer collects the first diary, reminds the respondent to complete the second one and sets a final meeting date to collect that diary.

If a person refuses to participate in the FES, the interviewer completes a non-interview report and informs the senior interviewer as soon as possible. Depending on the comments received, the senior interviewer will decide whether or not to start refusal conversion attempts. Note that sending a letter for refusals mentioning the importance of obtaining the data may be useful, but has its limits due to time constraints as the data must be gathered over a two week period in the interview month.

4.3 Monitoring and controls

All of the FES interviewers work under the supervision of a senior interviewer responsible for ensuring that the interviewers are very familiar with the survey's concepts and methods and for periodically overseeing the work and reviewing the documents completed. The senior interviewers work in turn under the supervision of program managers posted to each of Statistics Canada's regional offices.

At the end of each weekly reporting period, the interviewer returns to the respondent's residence to collect and review the diary from the previous week to ensure that it is complete and accurate. If the diary was not completed, the interviewer tries to complete it by asking the respondent to recall what his expenditures were. He also uses this opportunity to gather information on what factors may have influenced the respondent's habits and expenditures. Some of the factors might be the absence of members of the household during the week, the number of meals served to guests or other people not part of the household, the number of meals received free, donations of food, or food from the respondent's farm business or garden, or from hunting or fishing.

4.4 Non-response to the FES

Despite the interviewers' best efforts to gather the information, there are always a certain number of non-respondent households. It may not have been possible to contact them, an interview may not have been conducted for reasons beyond anyone's control, or the members of the household may have refused to participate in the survey. Collection non-response rates for the past two surveys are shown in Table 4.1.

Table 4.1
FES non-response rates

Reference year	Non-response rates			
	At collection time			Final total
	Total	No contact	Refusal	
1996	21.8	4.6	17.2	23.7
2001	28.5	11.4	17.0	28.5

In 2001, the final non-response rate is equal to that of the collection non-response rate. In 1996, the final non-response rate was 23.7% as a result of 280 questionnaires that were answered but could not be used. This difference is explained by a change in data processing between 1996 and 2001.

There are also respondents who report only one week of data. In 2001, for the 10 provinces, this situation occurred in slightly less than 5% of cases. In 1996, it was slightly less than 2%.

5. Data processing

The main steps in the processing of FES data are response coding, data capture, editing, imputation of partial non-response, and weighting. The latter will be covered in Section 6.

5.1 Coding, data capture and editing

Coding is an important step in the FES. Basic coding is done on the questionnaire and a lot of coding is done on the diaries: all the descriptions of the bought items declared in the diaries must be coded. Note that in 2001, the coding was done by specialized personnel from this survey, which was not the case in 1996.

The coding is done at the Head Office of Statistics Canada (SC). In the case of the diaries, the coders use software developed in the Income Statistics Division of STC to help choose the appropriate food codes and write the codes in red directly on the diaries. When the description of the item purchased is missing or not detailed enough, the most appropriate unspecified food code is given and will be handled in the automated imputation process mentioned in the next section.

In the case of food purchased from stores, each entry of a diary consists of the following fields: the description of the item purchased, the type of processing, the type of store, the number of units purchased, the net weight or volume per unit and the total cost of the purchase. All fields that are missing, other than the description of the item, are imputed by the coder using the other information present in the diary. For quality control, some respondents' questionnaires and diaries are selected randomly and checked by the senior coders.

After the coding is done, the questionnaires and diaries are captured. The main automated edit during capture is the one checking the validity of the food codes. If a food code is not valid, the keyer checks if it is a keying error. If this is not the case, the keyer enters a specific code so that when it goes through a series of automated edits after data capture, this food code will be flagged as needing correction.

In terms of data capture, a workload for a keyer is 20 respondents' questionnaires and diaries split into 4 batches of 5. As a quality check of data capture, a batch is selected randomly. If the data quality of the batch is acceptable, then the workload is accepted, if not then the 3 other batches are checked.

Finally, captured data pass through a series of automated edits. When there are errors, some of them are imputed automatically while the others, after a final look at the questionnaires and/or diaries, are imputed or flagged manually as needing to undergo the automated imputation process.

5.2 Imputation

As already mentioned in the previous section, during coding and data capture the questionnaires and diaries are submitted to some manual and automated edits, as well as to some manual or automated imputations including fields flagged as needing to go into the automated imputation process.

The fields that are flagged to be imputed are mostly missing fields. An example of a missing field on the questionnaire (FE2) is having an expenditure for restaurants but the number of meals purchased in them missing. An example of a missing field on the diary (FE3) is having beef as a purchased item, but the type of beef (ground, stewing, etc.) missing. More than one automated method of imputation is used, but the basic one in 2001 is the nearest-neighbour method.

This technique involves forming groups of similar households based on certain classifications (e.g., province of residence, quarter of data collection). Within those groups, each household requiring imputation (recipient) is matched to a household that has a complete questionnaire (donor) and resembles the other most closely with respect to certain characteristics (e.g. income, household size, etc.). The donor's data are imputed to the recipient as long as they satisfy the edit requirement for consistency with the data reported by the recipient. For example, if the recipient has indicated the purchase of prepared food then the imputed expenditures must be greater than zero.

If a donor cannot be found for some recipients during this first step, the donor imputation method is used again, dropping the household income resemblance, which is considered less significant than others. Many donor imputation steps are required in order to impute all recipients. For a few cases the most significant condition, which is the quarter, is also dropped to find a donor. For more details on the different donor imputation steps, see Reference [2].

In the case of the questionnaire (FE2), two sections are imputed: *Spending habits* (Section B) and *Food and beverages while away from home overnight or*

longer during the previous month (Section C). These sections are imputed independently since they are considered fairly independent from each other. Section B uses the nearest neighbour approach, but Section C in some cases uses other methods due to the difficulty of finding an appropriate donor. For example, some respondents have reported expenditures in restaurants but not the number of meals purchased or vice versa. These cases are imputed using the mean imputation method. The imputed mean used is the one at the national level.

In the case of the diary (FE3), for each day only the section *Food and beverages purchased from stores* is imputed. To impute this section, independent imputation runs are carried out based on the grouping of missing information needing imputation. Generally the imputation is done at the entry level but in some cases it is done at the day level. These cases occur when the respondent has only listed the total expenditures on food purchased from stores on a given day.

For imputation at the entry level, the respondent must have provided some additional information. For example, the respondent has purchased beef without specifying which kind of beef: ground, stewing, etc. In this case, the unit price is considered the most important matching criteria in finding a donor. If the unit price is missing then total expenditure is used for matching, along with other variables.

Some descriptions of items are broad, such as Vegetables for example. For the first time, in 2001, these descriptions are coded and imputed. Sometimes a donor cannot be found for some recipients because of a high expenditure reported with the description. To impute them, we sum for each potential donor all the expenditures in a day with the appropriate food entries (all vegetables for example). When a donor is found with a sum of expenditures matching the recipient's high expenditure, then all the appropriate food entries of the donor are assigned to the recipient.

Day level information will be imputed if the respondent has just listed the total expenditures on food in any day. The recipient day's expenditure is matched to the donor total day's expenditure. Days are considered as donors only if they do not have any imputation at the entry level. All entries of a selected donor's day are assigned to the recipient record. The recipient total expenditures will be overwritten with the donor total expenditures when the two numbers do not match exactly.

Finally, note that households providing data for only one week, out of two, are not imputed for the week they did not report.

6. Weighting and estimation

Estimates are based on the principle that each household in the sample represents a certain number of households in the target population, as defined in Section 2. Therefore each respondent household is assigned a survey weight,

which indicates how many households in the population it represents. This survey weight is generally the product of three factors: a design weight that incorporates data from the sample design, an adjustment factor for non-response that offsets non-respondent households, and finally an adjustment factor to re-adjust the estimates of some characteristics of the sample based on estimates coming from external sources considered more reliable.

6.1 Design weight

A household's design weight corresponds to the inverse of the probability that it is included in the sample. Since the FES is a probability survey, each household of the target population has a known probability of being selected in the sample. If, for example, a household's probability of being selected is 1 in 200, then this household will have a weight of 200.

The sample design, defining among other things the sample allocation method, determines the design weight. The FES uses the sample design of the LFS, which is a sample design self-weighted in each stratum. This means that the design weights established when the design is developed are equal in each of the stratum. To the extent that the sample design and sample allocation are not modified, the initial weights could be used. However, the stabilization steps and sub-sampling described in Section 3.3 alter the initial selection probabilities. Consequently, the LFS design weights are adjusted to take these changes into account.

Since the FES sample is a sub-group of the six LFS rotation groups, the survey's design weights are obtained by adjusting the LFS design weights based on the number of rotation groups used. This latter factor can vary from one stratum to another since the number of rotation groups used in each stratum is set so as to meet the specific needs of the FES for sample allocation. In most cases, because of the very small size of the FES, it was necessary to sub-sample within an LFS rotation group. The weights are therefore also adjusted to take this step into account.

6.2 Adjustment for total non-response

When a respondent fails to respond to certain questions on the FE2 questionnaire or leaves out certain information on the expenditures reported in the diary, the missing data are imputed using the methods described in Section 5.2. The weights are adjusted to compensate for total non-response (when the household cannot be contacted or refuses to respond).

The adjustment of the weights for total non-response is based on the hypothesis that the characteristics of the non-respondent households are similar to those of the respondent households and, therefore, the latter can be used to represent all households: respondents and non-respondents. To make the adjustment, the sample is first subdivided into non-response classes defined in such a way as to increase the chances that respondents and non-respondents have similar characteristics. These characteristics are chosen based on what the survey is looking to measure.

Non-response classes were defined quarterly since the size of the sample did not allow definition of classes for each month.

Non-response classes by quarter correspond to the different levels of urbanization in each region. The levels of urbanization are generally as follows: main metropolitan areas, urban areas of over one million inhabitants, other smaller urban areas and finally, rural areas. In some regions, it may be necessary to combine some levels because the sample is too small. In each of the three cities of the territories, there are only four non-response classes, each representing one quarter.

High-income household strata also constitute specific non-response classes in each of the regions in which they exist. Non-response classes do not overlap and when combined, they cover all of the target population.

For each non-response class, a non-response compensation factor is calculated and defined as the inverse of the weighted response rate of the class. This factor, in other words, is the ratio of the number of weighted sampled households, so that they represent all the households in the class, over the number of weighted respondent households. To avoid adjustment factors for non-response that are too high, as many non-response classes as possible are combined when the adjustment factor is higher than 2.

6.3 Adjustment to auxiliary data

In principle, by multiplying the design weight by the adjustment factor for non-response, it is possible to produce estimates at the population level. However, auxiliary data on the target population can be used to improve the survey's estimates. More reliable estimates can be produced if these auxiliary data are correlated to the main characteristics measured by the survey. For example, food expenditure is correlated to the size of the household. An incorrect distribution of the sample in terms of the size of the household would impact expenditure estimates. Using auxiliary data on the number of households by size, the weights can be adjusted to obtain a better distribution of the number of households.

In the FES, several sources of auxiliary data are used to adjust weights. First of all, post-censal estimates of population, produced by Statistics Canada's Demography Division, provide counts of the number of people by various age groups and by gender per province or territory. These counts correspond to population estimates at a given time based on census data and on information from administrative files, such as births, deaths, immigration, emigration, etc.

These demographic estimates are adjusted to take into account certain specific exclusions to household surveys, such as people living in institutions. From a design standpoint, they differ slightly from the FES target population by including people residing in collective households that are not institutions, for example, members of groups living in communal colonies and individuals residing full-time in hotels or rooming houses. However, this difference is considered negligible because these individuals represent less than 0.4% of the Canadian population.

These counts are then adjusted to reflect the FES target population at the end of the survey's reference year. This adjustment includes the exclusion of remote areas because they are not covered by the survey. Counts for six age groups for each region[6] are used to adjust the weights. The demographic counts of the number of people aged 18 or older and the number of people under 18 years of age in certain metropolitan areas[7] are also used.

The estimates of the number of households by size (one, two, or three or more persons) for each region and the number of households by certain types of household are also used to adjust the representativeness of the sample in these groups. For types of household, we use more specifically the number of households consisting of a lone-parent family and the number of households composed of parents with children who have never married.

For the city of Whitehorse, we use estimates, derived[8] from those of the Yukon, of the number of households by size (one, two, or three or more people). For the cities of Yellowknife and Iqaluit, we use only the estimate of the number of households. In the case of Yellowknife, the estimate was derived from that of the Northwest Territories. For Iqaluit, we used the preliminary estimate of the number of households following the 2001 Census given the unusual growth of this city as a result of the creation of the territory of Nunavut in 1999.

The adjustment of the weights to take into account all of the above counts is done simultaneously using a variant of the generalized regression estimator (GREG), which is based on the weighting method proposed by Lemaître and Dufour [3]. With this method, it is possible to match the survey's estimates with estimates from auxiliary sources and ensure, at the same time, that all members of the household will still have the same weight after adjustment. The adjustment factor calculated by the GREG is then applied to the design weight and to the non-response adjustment factor to produce the household's final weight.

A slight variation on this method had to be used for Yellowknife and Iqaluit. Although only the number of households was controlled, estimates of the number of households by size were unacceptable after application of the method as such. For further details, see Reference [4].

6.4 Other adjustments

During the in-person interviews with the questionnaire (FE2), respondents were asked to indicate the household expenditure for food purchased at stores over a four-week period. The expenditure estimate obtained from the questionnaire (FE2) was 8.3% higher than that obtained from the diary data (FE3). It is felt that the burden imposed by requiring the respondent to fill out the two diaries (FE3), which consists mainly of reporting a wide variety of items over a two-week

6. In Whitehorse, only two groups are used: under 18 and 18 or over. In Yellowknife and Iqaluit, there is no adjustment done in terms of number of people.
7. The metropolitan areas are as follows: St. John's, Halifax, Saint John, Québec, Montréal, Ottawa, Toronto, Winnipeg, Regina, Saskatoon, Calgary, Edmonton, Vancouver and Victoria.
8. Estimates for a city in one of the territories are obtained from equivalent estimates of its territory multiplied by the proportion of the territory that this city represents based on the 1996 Census.

period, may lead to respondent fatigue and may affect the quality of the responses obtained. Based on the assumption that the estimate of food purchased in stores from the questionnaire is more reliable than the corresponding diary estimate[9] due to the burden related to the diary, the values for each household of food purchased in stores locally and while on one-day trips were multiplied by 1.083.

To take into account the seasonality of food expenditures, the data collection is done during all 12 months of the year. The strategy chosen to realize this is to divide the sample evenly into 12 subsets – that is, one subset per month. Another potential adjustment to this type of survey would be a temporal adjustment. Data collection problems encountered in a specific month, for example, might justify this type of adjustment. To date, with the current adjustments, a temporal adjustment has not been considered necessary.

6.5 Estimates

The FES gathers food expenditure data from respondents over a two-week period. Some respondents only provide expenditures for one week. In order to calculate the estimate of average weekly expenditure per household for an item, we calculate the estimate of its total expenditure divided by the estimate of the number of diaries. As well, to calculate the percentage reporting for an item, we calculate the estimate of the number of diaries reporting an expenditure for that item divided by the estimate of the number of diaries.

7. Estimation of the sampling error

After the estimates have been calculated, it is necessary to determine the reliability of those estimates, or in other words, to estimate the sampling error associated with each estimate. The standard error or coefficient of variation (CV), which is simply the standard error expressed as a percentage of the estimate, is the common measure of the sampling error. It corresponds to the degree of variation observed in estimates after selecting a particular sample rather than another. Since the FES is a probability survey, it is possible to estimate the standard error of the estimates.

In the FES, the jackknife method is used to estimate the standard error. This method involves creating replicates of the sample based on FES data. As many replicates are created as there are first-stage sampling units (primary sampling unit or PSU)[10] by removing one PSU from the sample for each successive replicate. Each PSU is part of a stratum and when one PSU is removed, the design weights of the other PSUs in the stratum are adjusted to compensate for the removal. The final estimates are then recalculated using the adjustments to the auxiliary data described in Section 6.3 applied to the replicates. By repeating this operation for each PSU in the sample, we get the same number of estimates as there are PSUs. The variability between these estimates is used to estimate

9. Other analysis supports this assumption. For further details, see Reference [5].
10. The PSUs, in the case of the FES, are the clusters. For further details, see Section 3.

the standard error of the estimate from the whole sample. The detailed formula is found in Appendix A

.

7.1 Model for deriving an approximation of the CV for domain estimates

For operational reasons, it is not possible to produce CVs for all the various levels of aggregation that may be of interest to users, for example, by type of household, by level of urbanization or for certain metropolitan areas. The approach suggested to FES users is to calculate an approximation of the CV using a relationship of the number of diaries in which expenditures for an item were reported to the CV at an aggregate level (generally at the national level). This relationship, based on the CV's tendency to grow inversely proportional to the square root of the number of diaries with a non-zero amount, is shown in Appendix B.

7.2 Model for deriving an approximation of the CV from microdata files

Users of microdata files can use another approach to derive an approximation of the CV of an estimate, which will generally be more accurate than the one described above. This approach, which is relatively simply to apply, is described in greater detail in Reference [6]. It can only be used with the microdata files because both the data and the weights for each household are needed to calculate this approximation.

8. Suppression of data and confidentiality

A number of measures are taken to ensure that the estimates produced from the FES are sufficiently reliable to be published and that the anonymity of the households is respected.

8.1 Suppression of unreliable data in the tables of estimates

Since the coefficient of variation is an indicator of the reliability of the data, ideally we would use it to determine whether the estimates should be published or not. Estimates with a CV estimated at more than 33% are not reliable enough to be published.

However, the large number of estimates produced from the FES means that it is not possible to calculate the CVs for each of these estimates. A study conducted using the data of the 1996 and 2001 surveys revealed that the CVs are generally 33% when the number of diaries with a non-zero amount is around 40. This rule is therefore applied to determine whether the FES estimates can be published. Since this is an approximate rule, some estimates will be published even when the CV is higher than 33%, while some will not be published even though the CV is less than 33%. The report on the quality of the 2001 data [7] contains an evaluation of the performance of this rule.

It is important to note that even though estimates for a certain category of food are not released because fewer than 40 diaries contained a non-zero amount for that category, the data are retained in the estimates of the aggregate components.

8.2 Confidentiality of the microdata files

Although public use microdata files are developed from the data gathered by the FES, they are different from the files used by Statistics Canada for the dissemination of estimates. These differences are due mainly to a series of measures taken to protect the anonymity of the households responding to the survey.

9. Changes in the survey methodology

The survey's methodology remained similar to that of 1996 except for the strategies for imputation and for adjustment of the auxiliary data in the weighting. The sample design was also modified as a result of the reduction in the size of the sample for budget reasons. The goal at this time is to obtain estimates with reasonable variances at the level of the five Canadian regions rather than the ten provinces.

The weighting strategy was modified as part of a project to harmonize the auxiliary data adjustments in Statistics Canada surveys related to income statistics. For the FES, the principal changes were the use of more age groups for demographic counts and the addition of counts at the household type level. To ensure comparability of trend analyses, the 1996 FES was re-weighted using this new strategy and demographic estimates based on the 1996 Census rather than those based on the 1991 Census that were used in the original weighting.

Bibliography

[1] *Methodology of the Canadian Labour Force Survey*, Catalogue no. 71-526-XPB.

[2] Khan, K. (2002), *Imputation of the 2001 Food Expenditure Survey*, Household Survey Methods Division, Internal Document, Statistics Canada.

[3] Lemaître and Dufour (1987), *An Integrated Method for Weighting Persons and Families*, Survey Methodology, Vol.13, no. 2, pp.211-220, Statistics Canada.

[4] Arsenault, S. and Auger, S. (February 2003), *Pondération de l'Enquête sur les dépenses alimentaires 2001*, Household Survey Methods Division, Internal document, Statistics Canada.

[5] *FAMEX Reference Paper (1995)*, Labour and Household Surveys Analysis Division, Internal document, Statistics Canada.

[6] Beaumont, J.-F. (2000), *Variance Estimation For A Public Use Microdata File From A Complex Survey*, Household Survey Methods Division, Working Paper HSMD-2000-002F/A, Statistics Canada.

[7] Auger, S., Khan, K., Nadon, S. and Tremblay, J. (2005), *2001 Food Expenditure Survey – Data Quality Indicators*, Household Survey Methods Division, Catalogue no. 62F0026MIE-2005003, Statistics Canada.

[8] *Food Expenditure in Canada (2001)*, Catalogue no. 62-554-XIE.

APPENDIX A

Formula for calculating the variance of estimates using the jackknife method

In the jackknife procedure for estimating the variance of estimates, the variability between the estimates is measured using the following formula:

$$Var(\hat{Y}) = \sum_{h=1}^{H} \frac{n(h)-1}{n(h)} \sum_{i=1}^{n(h)} (\hat{Y} - \hat{Y}_{(hi)})^2$$

where

n(h) is the number of PSUs in stratum h

$\hat{Y}_{(hi)}$ is the estimate of Y when the PSU i of stratum h is removed.

The standard error is the square root of the variance.

APPENDIX B

Formula for approximating the CV for a domain (a sub-group of the population)

If CV(Y) represents the CV for the estimate of the average weekly expenditure per household of a certain characteristic for the whole of the population, then it is possible to calculate an approximation of the CV of the estimate of that characteristic for a domain (which can be considered a sub-group of the population, such as type of household, level of urbanization, ...) using the following equation:

$$CV\ (Y_d) = CV\ (Y) \times \sqrt{\frac{n}{n_d}}$$

where

n = *the number of diaries with a non-zero amount for the characteristic*

= $\dfrac{\text{the number of diaries} * \text{the estimate of the percentage reporting}^{11}}{100}$

n_d = *the number of diaries with a non-zero amount for the characteristic in domain d*

= $\dfrac{\text{the number of diaries in domain d} * \text{the estimate of the percentage reporting in domain d}}{100}$

In general, the CV, the number of diaries and the percentage reporting at the national level are used to calculate approximations for various domains. In a case where we want to calculate an approximation of the CV of a domain wholly contained in a single region (e.g. metropolitan area), it is preferable to use these values at the regional level since the regional CVs are published for the FES.

You will find an example of the calculation on the next page.

11. Percentage of diaries with a non-zero amount for the characteristic.

Calculation example

In Table 2b of the publication of the 2001 FES [8], there are different estimates by income group, but the CVs associated with these estimates are not found in the publication.

The CV of average weekly expenditure per household for food purchased in stores in the case of all households in the lowest income group, based on the formula provided above and using Tables 2a, 2c and the technical table 2b, is

$= 1.04\% \times \sqrt{10140/1756}$

$= 1.04\% \times 2.40$

$= 2.50\%$

where n is the number of diaries with a non-zero amount

= (number of diaries x percentage reporting) / 100

= (11,034 x 91.9) / 100

= 10,140

and n_d is the number of diaries from households in the lowest income group with a non-zero amount

= (number of diaries from households in the lowest income group x the percentage reporting for households in the lowest income group) / 100

= (2,011 x 87.3) / 100

= 1,756.

N° 62F0026MIF au catalogue — N° 004
ISSN : 1708-8887
ISBN : 0-662-74522-1

Document de recherche

Série de documents de recherche sur les dépenses des ménages

La méthodologie de l'Enquête sur les dépenses alimentaires
2001

par Sylvie Auger, Khushnood Khan et Johanne Tremblay

Division des méthodes d'enquêtes auprès des ménages
Immeuble R.H. Coats, Ottawa, K1A 0T6

Téléphone: 613 951-7355

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 (613) 951-7355; (888) 297-7355: revenu@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements concernant le Programme des services de dépôt	1 800 700-1033
Télécopieur pour le Programme des services de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements pour accéder au produit

Le produit nº 62F0026MIF au catalogue est disponible gratuitement. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Nos produits et services.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136. Les normes de service sont aussi publiées dans le site www.statcan.ca sous À propos de Statistique Canada > Offrir des services aux Canadiens.

Statistique Canada
Division de la statistique du revenu

Série de documents de recherche sur les dépenses des ménages

La méthodologie de l'Enquête sur les dépenses alimentaires
2001

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2005

Tous droits réservés. Le contenu de la présente publication peut être reproduit, en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux, et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire quelque contenu de la présente publication, ou de l'emmagasiner dans un système de recouvrement, ou de le transmettre sous quelque forme et par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Juillet 2005

N° 62F0026MIF2005004 au catalogue

ISSN : 1708-8887
ISBN : 0-662-74522-1

Périodicité : hors série

Ottawa

This publication is available in English upon request (Catalogue no. 62F0026MIE).

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Résumé

L'Enquête sur les dépenses alimentaires (EDA) est une enquête périodique recueillant des informations auprès des ménages sur les habitudes de dépenses alimentaires. La collecte des renseignements se fait principalement à partir de journaux hebdomadaires des achats que les répondants doivent remplir quotidiennement durant deux semaines consécutives.

Ce document présente une description détaillée de la méthodologie de cette enquête. On y décrit d'abord brièvement le plan d'échantillonnage qui est principalement basé sur le plan de l'Enquête sur la population active. Par la suite, on y présente les méthodes de collecte, de traitement des données, de pondération, d'estimation de la variance ainsi que de suppression des données non fiables dans les tableaux d'estimations.

Table des matières

1. Introduction 6

2. Population cible 6

3. Plan d'échantillonnage 7

3.1 Taille et répartition de l'échantillon de l'EDA 8
3.2 Plan d'échantillonnage de l'EPA (sélection des grappes) 9
3.3 Sélection de l'échantillon de l'EDA 11

4. Collecte des données 12

4.1 Méthode de collecte des données 12
4.2 Entrevue et procédures de suivi 13
4.3 Supervision et contrôles 13
4.4 Non-réponse à l'EDA 14

5. Traitement des données 15

5.1 Codage, saisie des données et vérification 15
5.2 Imputation 16

6. Pondération et estimation 17

6.1 Poids d'échantillonnage 18
6.2 Ajustement pour la non-réponse totale 18
6.3 Ajustement à l'information auxiliaire 19
6.4 Autres ajustements 21
6.5 Estimations 21

7. Estimation de l'erreur d'échantillonnage 22

7.1 Modèle pour dériver une approximation du CV pour les estimations des domaines 22
7.2 Modèle pour dériver une approximation du CV à partir des fichiers de microdonnées 22

8. Suppression de données et confidentialité 23

8.1 Suppression des données non fiables dans les tableaux d'estimations 23
8.2 Confidentialité des fichiers de microdonnées 23

9. Changements dans la méthodologie de l'enquête 23

Bibliographie 25

Annexe A 26

Formule pour le calcul de la variance des estimations à partir du jackknife

Annexe B 27

Formule d'approximation du CV pour un domaine (un sous-groupe de la population)

1. Introduction

L'enquête sur les dépenses alimentaires (EDA) est une enquête périodique qui a lieu à tous les 4 à 6 ans. Cette enquête recueille des informations auprès des ménages canadiens sur les habitudes de dépenses alimentaires.

Cette enquête sert à compléter l'Enquête sur les dépenses des ménages (EDM)[1] puisque l'EDA procure des renseignements détaillés sur les dépenses alimentaires qu'il n'est pas possible d'obtenir avec la méthodologie utilisée pour l'EDM. L'Enquête sur les dépenses alimentaires et l'Enquête sur les dépenses des ménages servent à mettre à jour les pondérations utilisées dans l'Indice des prix à la consommation.

Les données de l'EDA sont recueillies à partir d'entrevues personnelles au moyen d'un questionnaire papier (FE2), ainsi que par l'entremise de deux journaux hebdomadaires des achats (FE3) que le répondant doit remplir quotidiennement durant deux semaines consécutives. Ces données sont recueillies auprès d'un échantillon de ménages couvrant les dix provinces ainsi que les villes de Whitehorse et Yellowknife. En 2001, la ville d'Iqaluit, a été ajoutée. L'enquête se déroule tous les mois au cours d'une même année civile. Les données recueillies sont par la suite vérifiées et pondérées. Les produits de l'enquête comprennent la publication, les tableaux et les fichiers de microdonnées nécessaires aux divers utilisateurs.

L'objectif de ce document est de fournir une description détaillée de la méthodologie de l'enquête quant au plan d'échantillonnage, à la collecte et au traitement des données, à la production des estimations et des autres produits ainsi qu'aux règles régissant leur diffusion.

2. Population cible

La population cible de l'EDA comprend les individus vivant dans les ménages privés canadiens à l'exclusion des représentants officiels de pays étrangers qui vivent au Canada et leurs familles ainsi que les résidents des réserves indiennes et des terres publiques. La restriction aux ménages privés implique que les pensionnaires d'établissements institutionnels tels que les prisons, les hôpitaux pour malades chroniques, les résidences pour personnes âgées ainsi que les membres d'ordres religieux et d'autres groupes vivant en communauté, les membres des Forces Armées vivant dans les camps militaires et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres sont exclus.

1. L'EDM est une enquête annuelle où l'on demande à un échantillon de ménages canadiens de faire état de toutes les dépenses effectuées au cours d'une année civile. Les résultats de l'EDM sont publiés dans *Les habitudes de dépenses au Canada*, n° 62-202 au catalogue.

Nous n'avons pas recueilli de renseignements auprès des personnes qui vivent temporairement loin de leur famille (par exemple, les étudiants d'université) parce que nous obtenons l'information auprès de leur famille, si elle est choisie dans l'échantillon. Donc, si un ménage est constitué uniquement de personnes vivant temporairement loin de leur famille, ce ménage est considéré hors du champ de l'enquête.

De plus, en 2001, puisque la taille de l'échantillon est beaucoup plus petite que lors des enquêtes précédentes, les membres des ménages habitant des logements situés dans les régions éloignées sont exclus de l'enquête. Ces régions, en général peu peuplées, sont situées principalement dans la portion nordique de certaines provinces. Ces régions sont difficilement accessibles et par conséquent, la collecte de données auprès de ces ménages est très dispendieuse.

L'enquête couvre près de 97% de la population dans les dix provinces. Au Yukon, Territoires du Nord-Ouest et Nunavut, l'enquête ne couvre que les villes de Whitehorse, Yellowknife et Iqaluit ce qui représente 62%, 44% et 17% de leur population respective selon le Recensement de 1996.

3. Plan d'échantillonnage

Les données sur les dépenses alimentaires sont recueillies auprès de ménages provenant d'un échantillon de logements sélectionnés selon un plan d'échantillonnage stratifié à plusieurs degrés. Ce plan varie selon le niveau d'urbanisation, mais il s'agit généralement d'un échantillon à deux degrés, dont le premier correspond à un échantillon aréolaire, c'est-à-dire un échantillon de régions géographiques appelées grappes.

La sélection des grappes est effectuée pour l'ensemble de l'année. Cet ensemble de grappes sélectionnées est divisé également en 12 sous-ensembles, afin de permettre la collecte des données au cours des 12 mois de l'année, soit un sous-ensemble par mois. La collecte est effectuée durant toute l'année, afin de tenir compte du caractère saisonnier des dépenses alimentaires. La méthode utilisée pour obtenir les 12 sous-ensembles doit faire en sorte que les trois échantillons mensuels pour chaque trimestre couvrent l'ensemble du Canada.

Au deuxième degré, avant chaque mois de collecte, les logements sont sélectionnés à partir d'une liste de tous les logements privés des grappes du sous-ensemble retenu pour le mois concerné. Tous les logements sélectionnés où habitent des personnes de la population cible constituent l'échantillon de l'enquête.

Afin de réduire les coûts de collecte, l'EDA utilise dans une large mesure le même plan d'échantillonnage que l'Enquête sur la population active (EPA). Les logements de l'échantillon de l'EDA sont sélectionnés à partir des grappes de l'échantillon de l'EPA, mais les deux enquêtes utilisent des échantillons de logements différents. Les principaux aspects du plan d'échantillonnage en

grappes de l'EPA sont décrits à la section 3.2. Une description plus détaillée figure dans la publication faisant état de la méthodologie de l'EPA [1]. L'approche utilisée pour la sélection des logements de l'EDA, à partir des grappes de l'échantillon de l'EPA, est abordée à la section 3.3. La section suivante explique comment l'échantillon est réparti entre les cinq régions canadiennes et les trois villes des territoires.

3.1 Taille et répartition de l'échantillon de l'EDA

En raison de restrictions budgétaires, la taille de l'échantillon de l'enquête de 2001 a été réduite de façon considérable comparativement à l'enquête précédente de 1996. Lors de l'enquête de 1996, l'échantillon comptait 17 000 logements, soit environ 15 000 ménages, une fois exclus les logements vacants et les ménages hors du champ de l'enquête (par ex., les ménages constitués seulement d'individus ayant une résidence permanente ailleurs). Avec un tel échantillon, on s'attendait à ce que les estimations aux niveaux national et provincial, ainsi que celles pour les trois villes des territoires[2], soient fiables.

À l'étape de la planification de l'EDA de 2001, il a été décidé de réduire la taille de l'échantillon, pour la faire passer de 17 000 à 14 000 logements. En dépit de cette réduction, on s'attendait toujours à des estimations relativement fiables au niveau national, provincial et pour les trois villes des territoires. Quelques semaines après le début de la collecte des données, en janvier 2001, on a dû réduire la taille de l'échantillon pour la faire passer de 14 000 à 9 000 logements en raison de contraintes budgétaires. Étant donné que la collecte était déjà en cours, il n'a pas été possible de réduire l'échantillon de janvier. L'échantillon a été réduit de façon appropriée à partir du mois de février. On s'attendait dès lors à pouvoir produire des estimations fiables au niveau régional[3] plutôt qu'au niveau provincial.

L'échantillon total, à l'exclusion de l'échantillon des trois villes des territoires, a été réparti entre les cinq grandes régions canadiennes, afin d'obtenir un compromis entre la fiabilité au niveau national et celle au niveau régional pour le *Total des dépenses alimentaires hebdomadaires*. De façon plus particulière, la répartition de l'échantillon est établie en fonction de la variabilité des dépenses au titre des aliments dans chaque région et, dans une moindre mesure, du nombre de ménages au niveau régional.

Dans chaque région, l'échantillon a été réparti entre les provinces selon la proportion du nombre de ménages. Chaque échantillon provincial a été réparti en proportion directe avec la taille de la population pour les régions métropolitaines de recensement. La portion de l'échantillon allouée à l'extérieur des régions métropolitaines de recensement a suivi la répartition de l'EPA [1].

Le tableau 3.1 présente la taille de l'échantillon de chacune des provinces et des trois villes des territoires pour les enquêtes de 1996 et 2001. La taille de l'échantillon est présentée selon le nombre de logements et le nombre de

2. Dans les territoires, seules les villes Yellowknife, Whitehorse et Iqaluit sont couvertes par l'enquête.
3. Les cinq régions sont l'Atlantique, le Québec, l'Ontario, les Prairies et la Colombie-Britannique.

ménages (une fois exclus les logements vacants et les ménages hors du champ d'enquête).

Tableau 3.1
Taille de l'échantillon, selon la province et les trois villes des territoires

Région géographique	Taille de l'échantillon			
	1996		2001	
	Nombre de logements	Nombre de ménages	Nombre de logements	Nombre de ménages
Canada	**17 000**	**14 993**	**9 488**	**8 414**
Terre-Neuve-et-Labrador	691	592	233	199
Île-du-Prince-Édouard	621	516	79	67
Nouvelle-Écosse	1 347	1 143	396	329
Nouveau-Brunswick	1 124	941	304	261
Région de l'Atlantique	**3 783**	**3 192**	**1 012**	**856**
Région du Québec	**2 395**	**2 139**	**1 816**	**1 615**
Région de l'Ontario	**3 886**	**3 479**	**2 632**	**2 404**
Manitoba	1 023	895	451	387
Saskatchewan	1 369	1 211	393	335
Alberta	1 461	1 319	948	868
Région des Prairies	**3 853**	**3 425**	**1 792**	**1 590**
Région de la Colombie-Britannique	**2 212**	**2 038**	**1 558**	**1 422**
Total des régions	**16 129**	**14 273**	**8 810**	**7 887**
Whitehorse	435	363	250	202
Yellowknife	436	357	252	192
Iqaluit	0	0	176	133
Total des 3 villes des territoires	**871**	**720**	**678**	**527**

3.2 Plan d'échantillonnage de l'EPA (sélection des grappes)

Le plan d'échantillonnage de l'EPA est fondé sur les données du recensement et est remanié après chaque recensement décennal, afin de tenir compte des changements démographiques. Le plan d'échantillonnage actuel est fondé sur les données du Recensement de 1991.

Les principes qui sous-tendent le plan d'échantillonnage de l'EPA sont les mêmes pour chaque province. Tout d'abord, chaque province est divisée en un certain nombre de régions géographiques, constituées de régions économiques (RÉ), de régions économiques de l'assurance-emploi (RÉAE) et de leurs intersections. De façon plus particulière, chaque région métropolitaine de recensement constitue une région géographique, étant donné qu'il s'agit d'une RÉAE.

Chaque région géographique est divisée par la suite en types de secteurs, principalement des secteurs urbains, des secteurs ruraux et des régions éloignées. Le plan d'échantillonnage varie selon le type de secteur. Les régions éloignées représentent principalement la partie nord de certaines provinces, et

sont, pour la plupart, peu peuplées. Ces régions sont exclues de l'EDA. Le plan d'échantillonnage des secteurs urbains et ruraux est décrit ci-dessous.

Secteurs urbains

Dans certaines villes importantes comportant un grand nombre d'immeubles à appartements, on a eu recours à une liste des appartements et à une base aréolaire. Dans d'autres secteurs urbains, on a eu recours uniquement à une base aréolaire.

Une base aréolaire est formée d'une liste de régions géographiques. Ces régions sont combinées pour former des strates. On compte jusqu'à trois niveaux de stratification. Au niveau supérieur, l'objectif consiste généralement à constituer des strates compactes et contiguës au niveau géographique, tandis qu'au niveau inférieur, on vise la constitution de strates finales les plus homogènes possibles en ce qui a trait à certaines caractéristiques socio-économiques. Dans quelques grandes villes[4], des strates distinctes sont constituées à partir des secteurs de dénombrement où le revenu moyen des ménages est élevé (environ 100 000 $ ou plus).

Afin de réduire les coûts de la collecte, les logements qui constituent la strate finale ne sont pas sélectionnés directement. La strate est d'abord divisée en grappes. Dans les secteurs urbains, les grappes peuvent être des combinaisons de côtés d'îlot, de secteurs de dénombrement (SD) ou de parties de SD. Puis, les grappes sont sélectionnées (habituellement six, parfois un multiple de six) dans chaque strate, avec une probabilité proportionnelle à la taille de la grappe. Par exemple, si la taille d'une grappe est deux fois plus grande que celle d'une autre grappe, elle sera deux fois plus susceptible d'être sélectionnée que la deuxième.

La base de sondage d'appartements est une liste d'appartements dressée à partir de données de la Société canadienne d'hypothèques et de logement. Cette base de sondage fournit une meilleure représentation des personnes qui habitent en appartement et réduit les effets de l'augmentation des grappes découlant de la construction de nouveaux immeubles à appartements. Dans certaines villes[5], les strates des appartements sont divisées en deux catégories : à faible revenu (revenu moyen des ménages inférieur à 20 000 $) et régulière. Pour chaque strate de la base de sondage, les immeubles à appartements sont sélectionnés dans l'échantillon au premier degré, avec une probabilité proportionnelle au nombre d'appartements compris dans l'immeuble.

Dans les secteurs urbains à faible densité, à savoir principalement les villes très dispersées, on utilise un plan d'échantillonnage différent. L'échantillonnage se fait en trois étapes : tout d'abord, on sélectionne les villes à l'intérieur de la strate ; puis, les grappes (côtés d'îlot) sont choisies dans les villes ; enfin, les logements sont sélectionnés dans les grappes.

4. Montréal, Ottawa, Toronto, Hamilton, London, Winnipeg, Calgary et Vancouver.
5. Montréal, Ottawa-Hull, Toronto, Winnipeg, Calgary, Edmonton et Vancouver.

Secteurs ruraux

Dans les secteurs ruraux, on a uniquement recours à une base aréolaire. Des strates géographiques sont constituées en combinant deux ou trois divisions de recensement, qui sont par la suite subdivisées, lorsque leur nombre le permet, afin de constituer des strates qui sont homogènes du point de vue de certaines caractéristiques socio-économiques. Au premier degré d'échantillonnage, les secteurs de dénombrement sont sélectionnés à l'intérieur de chaque strate finale, selon une probabilité proportionnelle au nombre des logements compris dans le SD.

Dans les secteurs ruraux à faible densité, on a recours à une variante du plan d'échantillonnage. Deux ou trois unités primaires d'échantillonnage constituées d'un groupe de six SD sont sélectionnées au premier degré, puis un échantillon des logements est choisi à l'intérieur de chacun des SD des unités primaires d'échantillonnage sélectionnées.

3.3 Sélection de l'échantillon de l'EDA

Les intervieweurs visitent les grappes sélectionnées dans le plan d'échantillonnage de l'EPA et dressent une liste de tous les logements privés qu'elles comprennent. À partir de cette liste, on choisit un échantillon pour l'EPA et un autre pour l'EDA. Les logements sont sélectionnés grâce à un échantillonnage systématique.

Étant donné que l'EDA utilise un échantillon beaucoup plus restreint que l'EPA, les logements ne sont pas sélectionnés dans chaque grappe de l'EPA. L'EPA est une enquête par panel, dans laquelle les ménages demeurent dans l'échantillon pendant six mois. L'échantillon de l'EPA a été conçu afin de pouvoir être divisé en six sous-échantillons représentatifs, en vue de permettre le renouvellement d'un sixième de l'échantillon chaque mois. C'est pourquoi six grappes (ou un multiple de six) sont sélectionnées dans chaque strate finale, une par groupe de renouvellement. Cette méthode facilite la sélection d'un échantillon plus petit pour une autre enquête, étant donné que l'on peut utiliser un sous-ensemble des groupes de renouvellement. Il s'agit généralement de l'approche utilisée pour les enquêtes supplémentaires de l'EPA. Dans le cas de l'EDA, le nombre de groupes de renouvellement est déterminé au niveau de la strate, selon les besoins propres à l'enquête en ce qui a trait à la répartition provinciale et infraprovinciale de l'échantillon. Dans certains cas, seulement une partie d'un groupe de renouvellement est nécessaire. Lorsque tel est le cas, les logements sont supprimés de l'échantillon, grâce à une suppression systématique à partir d'une liste classée géographiquement.

L'échantillon des logements est obtenu après l'opération de listage des grappes. Étant donné que les taux d'échantillonnage sont déterminés au préalable, il peut exister une différence entre les tailles prévues et réelles des échantillons, dans les cas où le nombre de logements de la liste diffère du nombre utilisé pour l'élaboration du plan d'échantillonnage de l'enquête. Pour réduire les coûts de la collecte (étant donné que les tailles des grappes ont tendance à augmenter) et pour prévenir les disparités significatives dans les charges de travail des

intervieweurs, on a recours à deux méthodes pour contrôler la taille des échantillons.

La première méthode, qui est généralement utilisée, consiste à corriger le problème en supprimant systématiquement une partie des logements sélectionnés à l'origine. Le processus servant à maintenir la taille de l'échantillon au niveau souhaité est appelé stabilisation de l'échantillon. La deuxième méthode est utilisée lorsque le nombre de logements augmente de façon marquée dans certains secteurs urbains. Dans ce cas, on a plutôt recours au sous-échantillonnage des grappes. Il existe trois options, selon l'importance de l'augmentation et le niveau de similitude entre les nouveaux logements et ceux figurant dans la même strate : constituer des sous-grappes, créer une nouvelle strate, ou sous-échantillonner les logements de la grappe.

4. Collecte des données

L'EDA vise à recueillir des renseignements sur les dépenses alimentaires des ménages canadiens sur une base volontaire.

4.1 Méthode de collecte des données

La collecte des données est effectuée à partir d'entrevues personnelles au moyen d'un questionnaire papier (FE2), ainsi qu'avec deux journaux hebdomadaires des achats (FE3) que le répondant doit remplir quotidiennement durant deux semaines consécutives. L'enquête se déroule tous les mois au cours d'une même année civile et les données sont recueillies à partir des ménages provenant d'un échantillon différent de logements à chaque mois.

Le questionnaire sert essentiellement à recueillir des données sur certaines caractéristiques socio-économiques (par exemple, le nombre de personnes, le revenu du ménage), ainsi que des renseignements sur les habitudes d'achat des ménages et sur leurs dépenses alimentaires au cours du mois précédent.

Les journaux servent à recueillir les dépenses alimentaires quotidiennes aux magasins et aux restaurants. Dans le cas des dépenses effectuées aux magasins, les détails de ces dépenses comprennent la description de l'article acheté, le type de conditionnement (frais, congelé, en conserve, séché, autre), le type de magasin (supermarché, magasin spécialisé, dépanneur, autre), le nombre d'unités achetées, le poids net ou le volume par unité (en mesures métriques ou impériales) et le coût total de l'achat. Dans le cas des dépenses effectuées aux restaurants, les détails de ces dépenses comprennent le type de repas (déjeuner, dîner, souper, collation), le nombre de repas achetés, le type de restaurant (service aux tables, restauration rapide, cafétéria, autre) et le coût total de l'achat excluant les coûts, taxes et pourboires concernant les boissons alcoolisées.

En 2001, les répondants ont été fortement encouragés à joindre leurs reçus d'achat à la page du journal correspondant au jour de l'achat. Il y a eu aussi un autre élément nouveau en 2001, soit le carnet. Tous les membres du ménage

âgés de 10 ans et plus disposaient d'un petit carnet à apporter constamment avec eux pour y inscrire les produits au moment de l'achat.

Les membres d'un ménage

Pour obtenir les dépenses relatives à un ménage, il est essentiel de bien identifier les membres qui en font partie. La personne ou le groupe de personnes qui occupe un logement constitue un ménage. Pour l'EDA, les membres de ce ménage sont les personnes qui vivent dans le logement au moment de l'entrevue qui n'ont pas une résidence permanente ailleurs et ne sont pas membres d'un autre ménage.

4.2 Entrevue et procédures de suivi

Les entrevues sont menées par des intervieweurs de Statistique Canada dont certains sont également en charge de recueillir l'information de l'EPA. Ces intervieweurs reçoivent une formation particulière à l'EDA.

Une à deux semaines avant le mois de l'entrevue, le bureau régional envoie par la poste une lettre d'introduction aux occupants des logements sélectionnés soulignant l'importance de l'enquête et la confidentialité des renseignements recueillis ainsi qu'un dépliant donnant plus d'information concernant l'enquête. L'intervieweur se rend ensuite au logement. Lorsqu'il n'y a personne à la maison, d'autres tentatives sont effectuées auprès du ménage, par exemple en le visitant à différentes heures de la journée ou en tentant de le rejoindre par téléphone. Le numéro de téléphone, s'il est disponible, fait partie de l'information au sujet du logement qui est fournie à l'intervieweur. Si le numéro de téléphone est manquant, l'intervieweur essaiera de le retracer.

Lors du rendez-vous initial, l'intervieweur remplit le questionnaire (FE2) en transcrivant lui-même les réponses obtenues auprès du répondant. Par la suite, il remet les journaux (FE3) au répondant en donnant des indications sur la façon de bien remplir le journal et fixe un rendez-vous afin de revenir huit jours plus tard pour recueillir le journal complété. Lors de cette deuxième visite, l'intervieweur récupère le premier journal, rappelle au répondant de remplir le deuxième journal et fixe un dernier rendez-vous afin de récupérer celui-ci.

Si une personne refuse de participer à l'EDA, l'intervieweur complètera un rapport de non-interview et en informera l'intervieweur principal le plus tôt possible. Selon les commentaires fournis, l'intervieweur principal décidera s'il entame ou non les tentatives de conversion de refus. À noter qu'envoyer une lettre destinée aux refus mentionnant l'importance d'obtenir des données peut être utile mais a ses limites en raison de la contrainte de temps ; les données doivent être recueillies sur une période de deux semaines au cours du mois de l'entrevue.

4.3 Supervision et contrôles

Tous les intervieweurs de l'EDA travaillent sous la supervision d'un intervieweur principal qui est chargé de s'assurer à ce que les intervieweurs connaissent bien

les concepts et les méthodes de l'enquête ainsi que de contrôler périodiquement le travail et de revoir les documents remplis. Les intervieweurs principaux travaillent à leur tour sous la supervision des gestionnaires de programme qui sont postés à chacun des bureaux régionaux de Statistique Canada.

À la fin de chaque période de déclaration hebdomadaire, l'intervieweur retourne chez le répondant pour prendre et réviser le journal de la semaine précédente afin de s'assurer qu'il est complet et exact. Si le journal n'a pas été rempli, l'intervieweur tente de compléter le journal en demandant au répondant de se rappeler ses dépenses. Il en profite pour recueillir des renseignements sur les facteurs qui auraient pu influer sur les habitudes et les dépenses du répondant. Parmi ces renseignements, on retrouve l'absence de membres du ménage pendant la semaine, le nombre de repas servis à des invités ou à d'autres personnes ne faisant pas partie du ménage, le nombre de repas reçus gratuitement, les dons d'aliments ou les aliments provenant de l'exploitation agricole ou du jardin du répondant, ou les produits de la chasse ou de la pêche.

4.4 Non-réponse à l'EDA

Malgré toutes les tentatives effectuées par les intervieweurs pour obtenir l'information, il reste toujours un certain nombre de ménages non répondants. Par exemple, il a été impossible de les contacter ; une entrevue n'a pas été menée dû à des circonstances incontrôlables ; ou parce que les membres du ménage ont refusé de participer à l'enquête. Les taux de non-réponse à la collecte observés au cours des deux dernières enquêtes sont présentés au tableau 4.1.

Tableau 4.1
Taux de non-réponse à l'EDA

Année de référence	Taux de non-réponse			
	À la collecte			Total final
	Total	Pas de contact	Refus	
1996	21,8	4,6	17,2	23,7
2001	28,5	11,4	17,0	28,5

En 2001, le taux de non-réponse final est égal à celui de la collecte. En 1996, le taux de non-réponse final est 23,7 % en raison de 280 questionnaires répondus mais qui étaient inutilisables. Cette différence s'explique par un changement de traitement des données entre 1996 et 2001.

Il y a également les répondants qui ne rapportent qu'une semaine de données. En 2001, au niveau des dix provinces, il y en a eu un peu moins de 5 %. En 1996, ils étaient un peu moins de 2 %.

5. Traitement des données

Les principales étapes du traitement des données de l'EDA sont le codage des réponses, la saisie des données, la vérification, l'imputation pour la non-réponse partielle et la pondération. Cette dernière est abordée à la section 6.

5.1 Codage, saisie des données et vérification

Le codage constitue une étape importante de l'EDA. On procède à un codage de base du questionnaire, et à plusieurs opérations de codage des journaux : les descriptions de tous les articles achetés dont il est fait mention dans les journaux doivent être codées. À noter qu'en 2001, le codage a été effectué par le personnel spécialisé de cette enquête, ce qui n'était pas le cas en 1996.

Le codage est effectué au bureau central de Statistique Canada (SC). Dans le cas des journaux, les codeurs utilisent un logiciel élaboré à la Division de la statistique du revenu de SC, afin de les aider à choisir les codes d'aliments appropriés et inscrivent les codes en rouge directement dans les journaux. Lorsque la description de l'article acheté n'est pas fournie ou n'est pas suffisamment détaillée, on attribue le code d'aliment générique le plus approprié, qui par la suite sera imputé au moyen du processus d'imputation automatisée dont il est question dans la section suivante.

Dans le cas des aliments achetés au magasin, chaque entrée d'un journal est composée des champs suivants : la description de l'article acheté, le type de conditionnement, le type de magasin, le nombre d'unités achetées, le poids net ou le volume par unité et le coût total de l'achat. Tous les champs qui sont manquants, à l'exception de la description de l'article, sont imputés par le codeur, à partir des autres renseignements figurant dans le journal. Afin de contrôler la qualité, certains questionnaires et journaux des répondants sont sélectionnés de façon aléatoire et vérifiés par les codeurs principaux.

Une fois le codage effectué, on procède à la saisie des questionnaires et des journaux. La principale vérification automatisée qui est effectuée au moment de la saisie consiste à vérifier la validité des codes des aliments. Si un code d'aliment n'est pas valide, le responsable de la saisie des données vérifie s'il s'agit d'une erreur de saisie. Si ce n'est pas le cas, le responsable de la saisie entre un code spécifique qui fait en sorte que, lorsque la série de vérifications automatisées suivant la saisie des donnés aura été effectuée, le code d'aliment comportera une étiquette indiquant qu'il nécessite d'être corrigé.

Du point de vue de la saisie des données, la charge de travail d'un commis à la saisie est constituée de 20 questionnaires et journaux répartis en quatre lots de cinq. Afin de vérifier la qualité des données saisies, on sélectionne un lot de façon aléatoire. Si la qualité des données du lot est acceptable, l'ensemble des questionnaires est approuvé ; autrement, les trois autres lots sont vérifiés.

Enfin, les données saisies sont soumises à une série de vérifications automatisées. Lorsque des erreurs sont décelées, certaines font l'objet d'une imputation automatique tandis que les autres, après examen final des

questionnaires ou des journaux, sont imputées ou étiquetées manuellement pour indiquer qu'elles doivent être soumises au processus d'imputation automatisée.

5.2 Imputation

Comme il est mentionné dans la section précédente, au cours du codage et de la saisie des données, les questionnaires et les journaux sont soumis à certaines vérifications manuelles et automatisées, et font l'objet de certaines imputations manuelles ou automatisées, y compris l'étiquetage des champs qui doivent être soumis au processus d'imputation automatisée.

Les champs qui sont étiquetés en vue de l'imputation sont principalement manquants. Parmi les exemples de champs manquants dans le questionnaire (FE2) figurent les dépenses dans les restaurants sans indication du nombre de repas achetés. Parmi les exemples de champs manquants dans le journal (FE3) figure la déclaration d'un achat de bœuf, sans indication du type de bœuf (haché, en cubes, etc.). On a recours à plus d'une méthode d'imputation automatisée, mais la méthode principale utilisée en 2001 est celle du plus proche voisin.

Cette technique repose sur la constitution de groupes de ménages similaires, selon certaines classifications (par ex., province de résidence, trimestre de la collecte des données). À l'intérieur de ces groupes, chaque ménage qui doit être imputé (receveur) est apparié à un ménage qui a rempli un questionnaire (donneur) et qui s'apparente le plus à l'autre du point de vue de certaines caractéristiques (p. ex., revenu, taille du ménage, etc.). Les données du donneur sont imputées au receveur si ces données permettent de satisfaire les règles de vérification de cohérence avec les données déclarées par le receveur. Par exemple, si le receveur a indiqué avoir acheté des aliments préparés, les dépenses imputées doivent être supérieures à zéro.

Si on ne peut trouver de donneurs pour certains receveurs lors de cette première étape, on a recours encore une fois à la méthode d'imputation par donneur, en supprimant l'exigence liée à la similitude entre les revenus des ménages, qui est considérée comme moins importante que les autres caractéristiques. Plusieurs étapes d'imputation par donneur sont nécessaires afin de pouvoir imputer tous les receveurs. Dans quelques cas, la condition la plus importante, à savoir le trimestre, est aussi abandonnée pour trouver un donneur. Pour plus de détails concernant les différentes étapes d'imputation par donneur, voir la référence [2].

Dans le cas du questionnaire (FE2), deux sections sont imputées : *Habitudes de dépense du ménage* (section B) et *Aliments et boissons consommés hors du lieu de résidence – pendant une absence d'une nuit ou plus au cours du mois précédent* (section C). Ces sections sont imputées séparément, étant donné qu'elles sont considérées comme relativement indépendantes l'une de l'autre. Pour la section B, on utilise l'approche du plus proche voisin, mais pour la section C, on utilise d'autres méthodes dans certains cas, en raison de la difficulté de trouver un donneur approprié. Par exemple, certains répondants ont déclaré des dépenses au restaurant, et n'ont pas déclaré le nombre de repas achetés, ou vice versa. Ces cas sont imputés grâce à la méthode d'imputation par la moyenne. La moyenne utilisée pour imputer est celle au niveau national.

Dans le cas du journal (FE3), pour chaque jour, on impute uniquement la section des *Aliments et boissons achetés dans un magasin*. Afin d'imputer cette section, on procède à des imputations indépendantes selon le regroupement des données manquantes qui doivent être imputées. De façon générale, l'imputation est effectuée au niveau de l'entrée, mais dans certains cas, elle se fait au niveau de la journée. Ces cas surviennent lorsque le répondant a indiqué uniquement les dépenses totales au titre des aliments achetés dans un magasin pour une journée donnée.

Pour que l'imputation au niveau de l'entrée soit possible, le répondant doit avoir fourni certains renseignements additionnels. Par exemple, le répondant indique avoir acheté du bœuf, sans préciser le type : haché, en cubes, etc. Dans ce cas, le prix à l'unité est considéré comme le critère d'appariement le plus important pour trouver un donneur. Si le prix à l'unité est manquant, on utilise la dépense totale de même que d'autres variables pour apparier.

Certaines descriptions d'articles déclarés sont imprécises : *Légumes*, par exemple. Pour la première fois, en 2001, ces descriptions sont codées et imputées. Quelques fois, on ne peut pas trouver de donneurs pour certains receveurs, du fait d'une dépense élevée correspondant à la description. Pour l'imputation, on additionne pour chaque donneur possible toutes les dépenses d'une journée correspondant aux entrées appropriées (tous les légumes par exemple). Lorsqu'un donneur est trouvé avec le total des dépenses appropriées correspondant à la dépense élevée du receveur, on attribue toutes les entrées appropriées d'aliments du donneur au receveur.

Lorsque le répondant a indiqué uniquement les dépenses totales au titre des aliments pour une journée donnée, l'imputation est effectuée au niveau de la journée. On apparie les dépenses d'une journée du receveur à celles des donneurs. Les journées qui peuvent servir de donneurs sont celles qui ne comportent pas d'imputations au niveau des entrées sous-jacentes. Toutes les entrées pour une journée d'un donneur sélectionné sont attribuées au receveur. Les dépenses totales du receveur sont remplacées par les dépenses totales du donneur dans les cas où les deux chiffres ne correspondent pas exactement.

Enfin, il convient de souligner que les ménages qui n'ont fourni qu'une semaine de données sur deux ne font pas l'objet d'une imputation dans le cas de la semaine pour laquelle ils n'ont pas déclaré de données.

6. Pondération et estimation

Les estimations sont fondées sur le principe que chaque ménage de l'échantillon représente un certain nombre de ménages de la population cible, telle qu'elle a été définie dans la section 2. On attribue donc, à chaque ménage répondant, un poids d'enquête qui indique combien de ménages de la population sont représentés par ce ménage. Ce poids d'enquête est généralement le produit de trois facteurs : un poids d'échantillonnage qui incorpore les données du plan d'échantillonnage ; un facteur d'ajustement pour la non-réponse qui vise à compenser pour les ménages non-répondants et finalement un facteur

d'ajustement pour rajuster les estimations de certaines caractéristiques de l'échantillon en fonction d'estimations provenant de sources externes considérées plus fiables.

6.1 Poids d'échantillonnage

Le poids d'échantillonnage d'un ménage correspond à l'inverse de la probabilité qu'il soit inclus dans l'échantillon. Comme l'EDA est une enquête probabiliste, chaque ménage de la population cible a une probabilité connue d'être choisi dans l'échantillon. Si par exemple la probabilité de sélection d'un ménage est 1 sur 200, ce ménage aura un poids de 200.

Le plan d'échantillonnage définissant, entres autres, la méthode de répartition de l'échantillon, permet de déterminer le poids d'échantillonnage. L'EDA utilise le plan d'échantillonnage de l'EPA qui est un plan autopondéré dans chaque strate c'est-à-dire que les poids d'échantillonnage établis lors de l'élaboration du plan sont égaux dans chacune des strates. Dans la mesure où le plan d'échantillonnage et la répartition de l'échantillon ne sont pas modifiés, les poids initiaux pourraient être utilisés. Toutefois, les étapes de stabilisation et de sous-échantillonnage décrites dans la section 3.3 modifient les probabilités de sélection initiales. Les poids d'échantillonnage de l'EPA sont donc ajustés pour tenir compte de ces modifications.

Comme l'échantillon de l'EDA correspond à un sous-ensemble des 6 groupes de renouvellement de l'EPA, on détermine les poids d'échantillonnage de l'enquête en ajustant les poids d'échantillonnage de l'EPA en fonction du nombre de groupes de renouvellement retenus. Ce dernier facteur peut varier d'une strate à l'autre puisque le nombre de groupes de renouvellement retenu dans chaque strate est déterminé de sorte à satisfaire les besoins spécifiques à l'EDA en termes de répartition de l'échantillon. Étant donné que la taille de l'échantillon de l'EDA est très petite, on a eu la plupart du temps à sous-échantillonner dans un groupe de renouvellement de l'EPA. Les poids sont donc également ajustés pour tenir compte de cette étape.

6.2 Ajustement pour la non-réponse totale

Lorsque le répondant a omis de répondre à certaines questions du questionnaire FE2 ou qu'il a omis certaines informations concernant les dépenses rapportées sur le journal, les données manquantes sont imputées selon les méthodes décrites dans la section 5.2. Pour compenser la non-réponse totale, c'est-à-dire lorsqu'il a été impossible de contacter le ménage ou lorsque celui-ci a refusé de répondre, on ajuste les poids.

L'ajustement des poids pour la non-réponse totale est fondé sur l'hypothèse que les caractéristiques des ménages non-répondants sont similaires à celles des ménages répondants et, par conséquent, ces derniers peuvent être utilisés pour représenter l'ensemble des ménages : les répondants et les non-répondants. Afin de procéder à ce rajustement, l'échantillon est d'abord subdivisé en classes de non-réponse définies de sorte à augmenter les chances que les répondants et les non-répondants possèdent des caractéristiques semblables. Ces

caractéristiques sont choisies en fonction de ce que l'enquête cherche à mesurer.

Les classes de non-réponse ont été définies par trimestre car la taille de l'échantillon ne permettait pas de définir des classes pour chacun des mois.

Les classes de non-réponse par trimestre correspondent à différents niveaux d'urbanisation dans chaque région. Les niveaux d'urbanisation sont généralement les suivants : les principales régions métropolitaines, les régions urbaines de plus d'un million d'habitants, les autres régions urbaines plus petites et, finalement, les régions rurales. Dans certaines régions, il peut être nécessaire de regrouper certains niveaux parce que l'échantillon est trop petit. Dans chacune des trois villes des territoires, il n'y a que quatre classes de non-réponse, chacune représentant un trimestre.

Les strates de ménage à revenu élevé constituent également des classes de non-réponse particulières dans chacune des régions où elles existent. On notera que les classes de non-réponse ne se chevauchent pas et que, réunies, elles couvrent l'ensemble de la population cible.

Pour chaque classe de non-réponse, un facteur de compensation pour la non-réponse est calculé et défini comme étant l'inverse du taux de réponse pondéré de la classe. Ce facteur est, en d'autres mots, le rapport du nombre de ménages échantillonnés pondérés, afin qu'ils représentent tous les ménages de la classe, sur le nombre de ménages répondants pondérés. Pour éviter que les facteurs d'ajustement pour la non-réponse ne soient trop élevés, on regroupe autant que possible certaines classes de non-réponse lorsque le facteur d'ajustement est supérieur à 2.

6.3 Ajustement à l'information auxiliaire

En principe, en multipliant le poids d'échantillonnage par le facteur d'ajustement pour la non-réponse, on peut produire des estimations au niveau de la population. Toutefois, il est possible d'utiliser des données auxiliaires sur la population cible pour améliorer les estimations de l'enquête. Si ces données auxiliaires sont corrélées avec les principales caractéristiques mesurées par l'enquête, des estimations plus fiables peuvent être produites. Par exemple, les dépenses alimentaires sont corrélées avec la taille du ménage. Une mauvaise répartition de l'échantillon par rapport à la taille du ménage aurait un impact sur les estimations de dépenses. À partir de données auxiliaires sur le nombre de ménages selon la taille, on peut ajuster les poids de sorte à obtenir une meilleure distribution du nombre de ménages.

Dans l'EDA, on utilise plusieurs sources de données auxiliaires pour ajuster les poids. D'abord, les estimations post-censitaires de population, produites par la Division de la démographie de Statistique Canada, fournissent des comptes du nombre de personnes selon différents groupes d'âge et selon le sexe par province ou territoire. Ces comptes correspondent à des estimations de la population à une période donnée établies à partir des données du recensement et de l'information provenant de dossiers administratifs tels que les naissances, décès, immigration, émigration, etc.

Ces estimations démographiques sont ajustées pour tenir compte de certaines exclusions spécifiques aux enquêtes auprès des ménages comme les personnes vivant en institution. Conceptuellement, elles diffèrent légèrement de la population cible de l'EDA en incluant les personnes habitant les ménages collectifs qui ne sont pas des institutions, par exemple les membres de groupes vivant en communauté et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres. Cependant, on considère cette différence négligeable puisque ces individus représentent moins de 0,4 % de la population canadienne.

Ces comptes sont ensuite ajustés pour refléter la population cible de l'EDA à la fin de l'année de référence de l'enquête. Cet ajustement comprend l'exclusion des régions éloignées puisqu'elles ne sont pas couvertes par l'enquête. Des comptes pour six groupes d'âge pour chacune des régions[6] sont utilisés pour ajuster les poids. Les comptes démographiques du nombre de personnes de 18 ans ou plus et du nombre de personnes de moins de 18 ans dans certaines régions métropolitaines[7] sont également utilisés.

Les estimations du nombre de ménages selon la taille (une, deux ou trois personnes ou plus) pour chaque région ainsi que du nombre de ménages selon certains types de ménages sont également utilisées pour ajuster la représentativité de l'échantillon dans ces groupes. Pour les types de ménage, on utilise plus spécifiquement le nombre de ménages constitués d'une famille monoparentale et le nombre de ménages composés de parents avec des enfants jamais mariés.

Pour la ville de Whitehorse, on utilise les estimations, dérivées[8] à partir de celles du Yukon, du nombre de ménages selon la taille (une, deux ou trois personnes ou plus). Pour les villes de Yellowknife et Iqaluit, on utilise seulement l'estimation du nombre de ménages. Dans le cas de Yellowknife, l'estimation a été dérivée à partir de celle des Territoires du Nord-Ouest. Dans le cas d'Iqaluit, on a utilisé l'estimation préliminaire du nombre de ménages suite au recensement de 2001 en raison de l'expansion inhabituelle de cette ville suite à la création du territoire du Nunavut en 1999.

L'ajustement des poids pour tenir compte de tous les comptes décrits ci-dessus est effectué simultanément en utilisant une variante de l'estimateur de régression généralisé (ERG) fondée sur la méthode de pondération proposée par Lemaître et Dufour [3]. Cette méthode permet d'obtenir la concordance des estimations de l'enquête avec les estimations provenant des sources auxiliaires tout en assurant que tous les membres du ménage auront encore un poids identique après l'ajustement. Le facteur d'ajustement calculé par l'ERG est ensuite appliqué au

6. À Whitehorse, on utilise seulement deux groupes : moins de 18 ans et 18 ans ou plus. À Yellowknife et Iqaluit, il n'y a aucun ajustement de fait au niveau du nombre de personnes.
7. Il s'agit des régions métropolitaines suivantes : St. John's, Halifax, Saint John, Québec, Montréal, Ottawa, Toronto, Winnipeg, Regina, Saskatoon, Calgary, Edmonton, Vancouver et Victoria.
8. Les estimations d'une ville d'un territoire sont obtenues à partir des estimations équivalentes de son territoire multipliées par la proportion qu'occupait cette ville par rapport à celui-ci selon le recensement de 1996.

poids d'échantillonnage et au facteur d'ajustement pour la non-réponse pour produire le poids final du ménage.

Dans les cas de Yellowknife et Iqaluit, une légère variante à cette méthode a dû être apportée. Bien qu'on ne contrôlait que le nombre de ménages, les estimations du nombre de ménages par taille de ménage étaient inacceptables suite à l'application de la méthode telle quelle. Pour plus de détails, voir la référence [4].

6.4 Autres ajustements

Au cours de l'entrevue en personne sur le questionnaire (FE2), on a demandé aux répondants d'indiquer les dépenses du ménage pour les aliments achetés au magasin pendant une période de quatre semaines. L'estimation des dépenses obtenue par l'entremise du questionnaire (FE2) est supérieure de 8,3 % à celle obtenue avec les données des journaux (FE3). On considère que le fardeau imposé en demandant au répondant de remplir les deux journaux (FE3), qui est principalement de rapporter une grande variété d'articles sur une période de deux semaines, peut entraîner la fatigue du répondant et avoir un impact sur la qualité des réponses obtenues. Suivant l'hypothèse que l'estimation des aliments achetés au magasin selon le questionnaire est plus fiable que l'estimation correspondante du journal[9] en raison du fardeau associé au journal, les valeurs de chaque ménage pour les aliments achetés au magasin localement et à l'occasion de déplacements d'une seule journée ont été multipliées par 1.083.

Afin de tenir compte du caractère saisonnier des dépenses alimentaires, la collecte des données est effectuée au cours des 12 mois de l'année. La stratégie adoptée afin de réaliser ceci est de diviser l'échantillon en 12 sous-ensembles égaux, soit un sous-ensemble par mois. Un autre ajustement potentiel à ce genre d'enquête serait un ajustement temporel. Par exemple, des problèmes de collecte rencontrés pour un mois en particulier pourraient justifier ce genre d'ajustement. Jusqu'à maintenant, avec les ajustements déjà présents, un ajustement temporel n'a pas été jugé nécessaire.

6.5 Estimations

L'EDA recueille auprès des répondants leurs dépenses alimentaires sur une période de deux semaines. Il arrive que certains répondants ne fournissent des dépenses que pour une semaine. Afin de calculer l'estimation de la dépense hebdomadaire moyenne par ménage pour un article, on calcule l'estimation de sa dépense totale divisée par l'estimation du nombre de journaux. De même, afin de calculer le pourcentage répondant d'un article, on calcule l'estimation du nombre de journaux ayant rapportés une dépense pour l'article divisée par l'estimation du nombre de journaux.

9. D'autres analyses appuient cette hypothèse. Pour plus de détails, voir la référence [5].

7. Estimation de l'erreur d'échantillonnage

Après avoir calculé les estimations, il faut en déterminer la fiabilité ou, en d'autres mots, estimer l'erreur d'échantillonnage associée à chaque estimation. L'erreur-type ou le coefficient de variation (CV), qui est simplement l'erreur-type exprimée en tant que pourcentage de l'estimation, est la mesure courante de l'erreur d'échantillonnage. Elle correspond au degré de variation que l'on observe dans les estimations suite au choix d'un échantillon particulier plutôt qu'un autre. Comme l'EDA est une enquête probabiliste, on peut estimer l'erreur-type des estimations.

Dans l'EDA, on utilise l'estimateur jackknife pour estimer l'erreur-type. Cette méthode consiste à créer des répliques d'échantillon à partir des données de l'EDA. On crée autant de répliques qu'il y a d'unités primaires d'échantillonnage (UPE)[10] en enlevant à tour de rôle une UPE de l'échantillon. Chaque UPE fait partie d'une strate et lorsque l'UPE est enlevée, les poids d'échantillonnage des autres UPE de la strate sont ajustés pour compenser ce retrait. On recalcule ensuite les estimations finales en appliquant les ajustements aux données auxiliaires décrits à la section 6.3 aux répliques. En répétant cette opération pour chaque UPE de l'échantillon, on obtient autant d'estimations qu'il y a d'UPE. La variabilité entre ces estimations sert à estimer l'erreur-type de l'estimation provenant de l'échantillon complet. La formule détaillée se trouve à l'annexe A.

7.1 Modèle pour dériver une approximation du CV pour les estimations des domaines

Pour des raisons opérationnelles, il n'est pas possible de produire les CV pour tous les différents niveaux d'agrégation qui peuvent intéresser les utilisateurs, par exemple, par type de ménage, par niveau d'urbanisation ou pour certaines régions métropolitaines. L'approche qui est suggérée aux utilisateurs de l'EDA est de calculer une approximation du CV en utilisant une relation entre le nombre de journaux dans lesquels des dépenses pour un article ont été déclarées et le CV à un niveau agrégé (généralement au niveau national). Cette relation, basée sur la tendance du CV à croître proportionnellement à une diminution de la racine carrée du nombre de journaux ayant un montant différent de zéro, est illustrée dans l'annexe B.

7.2 Modèle pour dériver une approximation du CV à partir des fichiers de microdonnées

Les utilisateurs des fichiers de microdonnées peuvent se servir d'une autre approche pour dériver une approximation du CV d'une estimation, qui sera généralement plus performante que celle décrite ci-dessus. Cette approche, relativement simple à appliquer, est décrite plus en détails dans la référence [6]. Elle est utilisable seulement à partir du fichier de microdonnées puisqu'il est nécessaire d'avoir les données et les poids de chaque ménage pour calculer cette approximation.

10. Les UPE, dans le cas de l'EDA, sont les grappes. Pour plus de détails, voir la section 3.

8. Suppression de données et confidentialité

Certaines mesures sont prises pour assurer que les estimations produites à partir de l'EDA sont suffisamment fiables pour être publiées et que l'anonymat des ménages répondants est respecté.

8.1 Suppression des données non fiables dans les tableaux d'estimations

Comme le coefficient de variation est un indicateur de la fiabilité des données, idéalement on l'utiliserait pour déterminer si les estimations devraient être publiées ou non. Les estimations dont le CV est estimé à plus de 33% ne sont pas suffisamment fiables pour être publiées.

Par contre, le grand nombre d'estimations produites à partir de l'EDA fait en sorte qu'il n'est pas possible de calculer les CV pour chacune de ces estimations. Une étude effectuée à partir de données des enquêtes 1996 et 2001 a démontré que les CV atteignent en général 33% lorsque le nombre de journaux ayant un montant différent de zéro est autour de 40. Cette règle est donc appliquée pour déterminer si les estimations de l'EDA peuvent être publiées ou non. Comme il s'agit d'une règle approximative, certaines estimations seront publiées même si le CV est supérieur à 33% et certaines estimations ne seront pas publiées malgré un CV inférieur à 33%. Une évaluation de la performance de cette règle se trouvent dans le rapport sur la qualité des données de 2001 [7].

On doit noter que même si les estimations pour une certaine catégorie d'aliments ne sont pas diffusées parce que moins de 40 journaux ont un montant différent de zéro pour cette catégorie, les données sont retenues dans les estimations des composantes agrégées.

8.2 Confidentialité des fichiers de microdonnées

Quoique des fichiers de microdonnées à grande diffusion soient produits à partir des données recueillies par l'EDA, ceux-ci sont différents des fichiers utilisés par Statistique Canada pour la diffusion des estimations. Ces différences proviennent principalement d'une série de mesures prises pour protéger l'anonymat des ménages qui ont répondu à l'enquête.

9. Changements dans la méthodologie de l'enquête

La méthodologie de l'enquête est demeurée semblable à celle de 1996 à l'exception des stratégies d'imputation et d'ajustement aux données auxiliaires dans la pondération. Le plan d'échantillonnage a également été modifié suite à la diminution de la taille d'échantillon pour des raisons budgétaires. On a maintenant comme objectif d'obtenir des estimations avec des variances raisonnables au niveau des cinq grandes régions canadiennes au lieu des dix provinces.

La stratégie de pondération a été modifiée dans le cadre d'un projet qui vise à harmoniser les ajustements aux données auxiliaires dans les enquêtes de Statistique Canada liées à la statistique du revenu. Pour l'EDA, les principaux changements ont été l'utilisation de plus de groupes d'âge pour les comptes démographiques et l'ajout de comptes au niveau des types de ménages. Afin d'assurer la comparabilité dans les analyses des tendances, l'EDA 1996 a été repondérée à l'aide de cette nouvelle stratégie et d'estimations démographiques basées sur le recensement de 1996 au lieu de celles basées sur le recensement de 1991 utilisées lors de la pondération originale.

Bibliographie

[1] *Méthodologie de l'enquête sur la population active du Canada*, N° 71-526-XPB au catalogue.

[2] Khan, K. (2002), *Imputation of the 2001 Food Expenditure Survey*, Division des méthodes d'enquêtes auprès des ménages, Document interne, Statistique Canada.

[3] Lemaître et Dufour (1987), *Une méthode intégrée de pondération des personnes et des familles*, Technique d'enquête, Vol.13, n° 2, pp.211-220, Statistique Canada.

[4] Arsenault, S. et Auger, S. (Février 2003), *Pondération de l'Enquête sur les dépenses alimentaires 2001*, Division des méthodes d'enquêtes auprès des ménages, Document interne, Statistique Canada.

[5] *FAMEX Reference Paper (1995)*, Division de l'analyse des enquêtes auprès des ménages et sur le travail, Document interne, Statistique Canada.

[6] Beaumont, J.-F. (2000), *Estimation de variance pour un fichier de microdonnées à grande diffusion provenant d'une enquête complexe*, Division des méthodes d'enquêtes auprès des ménages, Document de travail HSMD-2000-002F/A, Statistique Canada.

[7] Auger, S., Khan, K., Nadon, S. et Tremblay, J. (2005), *Enquête sur les dépenses alimentaires de 2001 – Indicateurs de qualité des données*, N° 62F0026MIF-2005003 au catalogue, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.

[8] *Dépenses alimentaires au Canada (2001)*, N° 62-554-XIF au catalogue.

Annexe A

Formule pour le calcul de la variance des estimations à partir du jackknife

Dans la procédure jackknife pour estimer la variance des estimations, on mesure la variabilité entre les estimations à l'aide de la formule suivante :

$$Var(\hat{Y}) = \sum_{h=1}^{H} \frac{n(h)-1}{n(h)} \sum_{i=1}^{n(h)} (\hat{Y} - \hat{Y}_{(hi)})^2$$

où

n(h) est le nombre d'UPE dans la strate h

$\hat{Y}_{(hi)}$ est l'estimation de Y quand l'UPE i de la strate h est enlevée.

L'erreur-type est la racine carrée de la variance.

Annexe B

Formule d'approximation du CV pour un domaine (un sous-groupe de la population)

Si CV(Y) représente le CV pour l'estimation de la dépense hebdomadaire moyenne par ménage d'une certaine caractéristique pour toute la population, alors on peut calculer une approximation du CV de l'estimation de cette caractéristique pour un domaine (que l'on peut considérer comme un sous-groupe de la population tel qu'un type de ménage, un niveau d'urbanisation, ...) à partir de l'équation suivante :

$$CV\ (Y_d) = CV\ (Y) \times \sqrt{\frac{n}{n_d}}$$

où

n = *le nombre de journaux ayant un montant différent de zéro pour cette caractéristique*

= $\dfrac{\textit{le nombre de journaux} * \textit{l'estimation du pourcentage déclarant}^{11}}{100}$

n_d = *le nombre de journaux ayant un montant différent de zéro pour cette caractéristique dans le domaine d*

= $\dfrac{\textit{le nombre de journaux dans le domaine d} * \textit{l'estimation du pourcentage déclarant dans le domaine d}}{100}$

Généralement on utilise le CV, le nombre de journaux et le pourcentage déclarant au niveau national pour calculer les approximations pour les différents domaines. Dans le cas où on cherche à calculer une approximation du CV pour un domaine entièrement contenu dans une seule région (par ex. région métropolitaine), il est préférable d'utiliser ces valeurs à l'échelle régionale puisque les CV régionaux sont publiés pour l'EDA.

Vous trouverez un exemple de calcul à la page suivante.

11. Pourcentage de journaux contenant un montant différent de zéro pour cette caractéristique.

Exemple de calcul

Dans le tableau 2b de la publication de l'EDA 2001 [8], on retrouve différentes estimations selon les tranches de revenu, par contre les CVs associés à ces estimations ne se retrouvent pas dans la publication.

Le CV des dépenses hebdomadaires moyennes des ménages pour les aliments achetés au magasin dans le cas de tous les ménages de la tranche de revenu la plus basse, selon la formule ci-dessus et à l'aide des tableaux 2a, 2c et du tableau technique 2b, est

$= 1{,}04\% \times \sqrt{10140/1756}$

$= 1{,}04\% \times 2{,}40$

$= 2{,}50\%$

où n est le nombre de journaux ayant un montant différent de zéro

= (nombre de journaux x pourcentage déclarant) / 100

= (11 034 x 91,9) / 100

= 10 140

et n_d est le nombre de journaux des ménages de la tranche de revenu la plus basse ayant un montant différent de zéro

= (nombre de journaux des ménages de la tranche de revenu la plus basse x pourcentage déclarant pour les ménages de la tranche de revenu la plus basse) / 100

= (2 011 x 87,3) / 100

= 1 756.